

# Isolation and characterization of the first xylanolytic hyperthermophilic euryarchaeon *Thermococcus* sp. 2319x1 and its unusual multidomain glycosidase

Sergey N. Gavrilov<sup>1#</sup>, Christina Stracke<sup>2#</sup>, Kenneth Jensen<sup>3</sup>, Peter Menzel<sup>4</sup>, Verena Kallnik<sup>2</sup>, Alexei Slesarev<sup>1,5</sup>, Tatyana Sokolova<sup>1</sup>, Kseniya Zayulina<sup>1</sup>, Christopher Bräsen<sup>2</sup>, Elizaveta A. Bonch-Osmolovskaya<sup>1</sup>, Xu Peng<sup>4</sup>, Ilya V. Kublanov<sup>1\*</sup>, Bettina Siebers<sup>2\*</sup>

#Both authors contributed equally

\*Corresponding authors

<sup>1</sup>Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia

<sup>2</sup>Molecular Enzyme Technology and Biochemistry (MEB), Biofilm Centre, Centre for Water and Environmental Research (CWE), University Duisburg-Essen, Essen, Germany

<sup>3</sup>Novozymes A/S, Bagsværd, Denmark

<sup>4</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark

<sup>5</sup>Fidelity Systems, Inc., Gaithersburg, MD USA

## 1. Introduction

**Supplementary Table 1.** Characterized and predicted cellulases and xylanases from hyperthermophilic Archaea.

Family	GH3	GH5	GH9	GH10	GH11	GH12	GH16	GH26	GH43	GH51	GH94
Activities	beta-xylo-gluco	beta-xylo-many (mainly gluco endo)	beta-xylo-(mainly endo)	endo-xylo	endo-xylo	beta-xylo-gluco	beta-(mainly endo)	beta-arabino	xylo-arabino	beta-xylo-glucosidases	beta-chitotriosidases
# proteins from cultivated thermophilic archaea	<b>83</b>	<b>55</b>	<b>4</b>	<b>14</b>	<b>5</b>	<b>62</b>	<b>18</b>	<b>5</b>	<b>14</b>	<b>18</b>	<b>8</b>
<b>Crenarchaeota</b>											
<i>Acidianus hospitalis</i> W1							<b>1</b>				
<i>Acidilobus saccharovorans</i> 345-15	<b>1</b>	<b>1</b>					<b>3</b>				
<i>Caldivirga maquilingensis</i> IC-167	<b>2</b>	<b>1</b>					<b>5</b>				
<i>Ignisphaera aggregans</i> DSM 17230	<b>3</b>	<b>3</b>					<b>5</b>		<b>1</b>	<b>1</b>	
<i>Metallosphaera cuprina</i> Ar-4							<b>1</b>				
<i>Sulfolobus acidocaldarius</i> (all strains)							<b>4</b>				
<i>Sulfolobus islandicus</i> (all strains)	<b>10</b>	<b>10</b>					<b>20</b>				
<i>Sulfolobus solfataricus</i> (all strains)	<b>5</b>	<b>5</b>					<b>12</b>				
<i>Thermofilum carboxyditrophicus</i> 1505		<b>1</b>									
<i>Thermofilum pendens</i> Hrk 5	<b>1</b>	<b>1</b>					<b>1</b>				
<i>Thermofilum sp.</i> 1807-2											
<i>Thermofilum sp.</i> 1910b											
<i>Thermoproteus uzoniensis</i> 768-20		<b>1</b>					<b>2</b>				
<i>Thermoproteus tenax</i> Kra 1							<b>1</b>				
<i>Thermosphaera aggregans</i> DSM 11486							<b>1</b>				
<i>Vulcanisaeta distributa</i> DSM 14429		<b>1</b>					<b>1</b>				

---

## Euryarchaeota

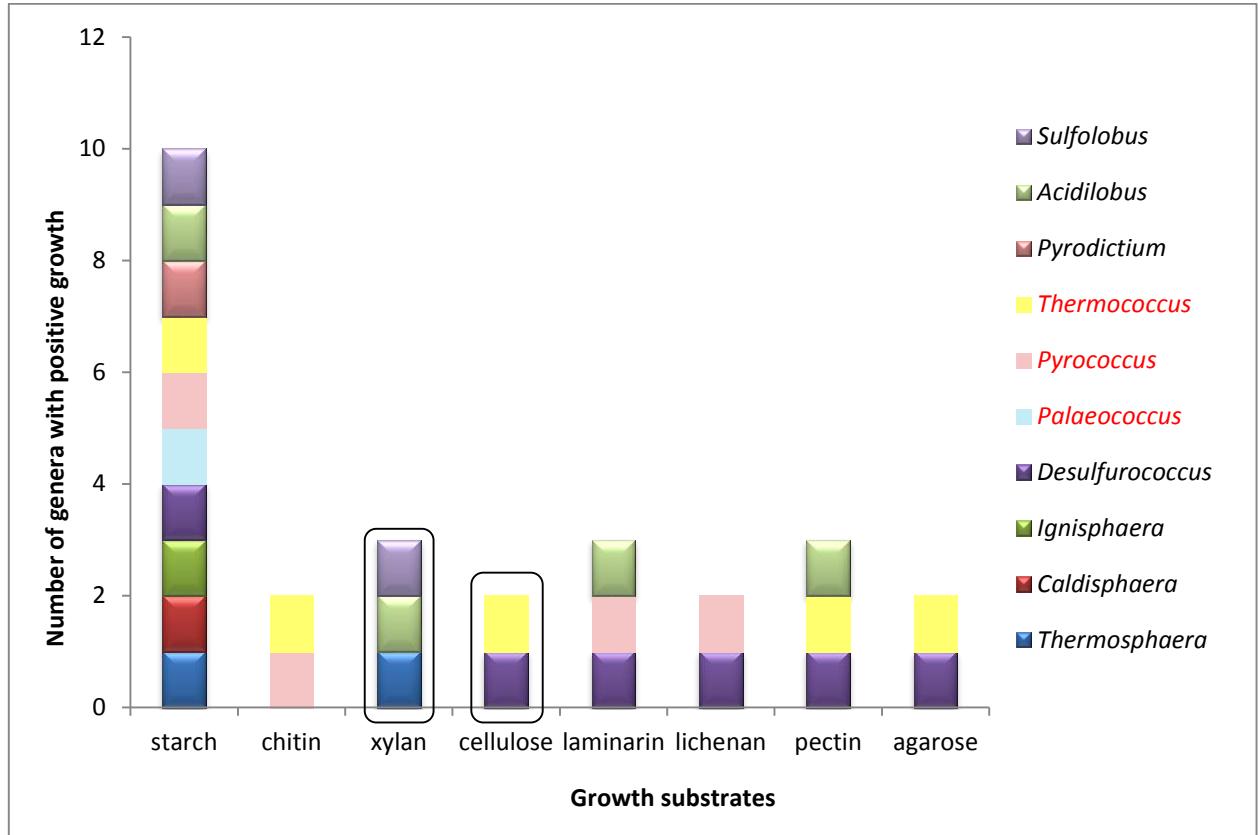
---

<i>Palaeococcus pacificus</i>		<b>1</b>		
DY20341				
<i>Picrophilus torridus</i> DSM 9790	<b>1</b>			
<i>Pyrococcus abyssi</i> GE5	<b>1</b>			
<i>Pyrococcus furiosus</i> COM1		<b>1</b>	<b>1</b>	
<i>Pyrococcus furiosus</i> DSM 3638		<b>1</b>	<b>1</b>	
<i>Pyrococcus horikoshii</i> OT3	<b>1</b>			
<i>Pyrococcus sp.</i> ST04	<b>1</b>			
<i>Thermococcus sp.</i> ES1	<b>1</b>			
<i>Thermococcus sp.</i> AM4		<b>1</b>		
<i>Thermococcus guaymasensis</i> DSM 11113		<b>1</b>		
<i>Thermococcus sibiricus</i> MM 739		<b>3*</b>	<b>1</b>	
<i>Thermoplasma volcanium</i> GSS1	<b>1</b>			

---

According to CAZy.

\* - very distant, not in CAZy



**Supplementary Figure 1.** Polysaccharide-degrading hyperthermophilic Archaea. Euryarchaeal genera are in red and their elements are in 2D; Crenarchaeal genera are in black and their elements are in 3D. Genera, growing on xylan or cellulose are boxed. Alpha-linked starch and its derivatives are widely utilized by hyperthermophilic archaea while beta-linked polysaccharides, such as cellulose and xylan are used rarely, mainly by representatives of the Crenarchaeota.

## 2. Material and Methods

### 2.1 Cloning, expression and purification of the MDG and truncated versions

**Supplementary Table 2.** Primers for In-Fusion® cloning of the *mdg* gene and its truncated versions. The 15 bp extensions complementary to the linearized expression vector pET24a are underlined.

Primer ID:	Sequence:
fwd- <i>mdg</i> (without signalpeptide)	5'-TCGCGGATCCGAATTATGACGACTGACACAAGCACG-3'
rvs- <i>mdg</i> (GH5-12-12-CBM2-2)	5'- <u>G</u> TGCTCGAGTGCGGCAAGGATTGACTCCAAGGAG-3'
rvs- <i>mdg-trunc1</i> (GH5-12-12)	5'- <u>G</u> GTGCTCGAGTGCGGGAGCAGCTTAGAAC-3'
rvs- <i>mdg-trunc2</i> (GH5-12)	5'- <u>G</u> GTGCTCGAGTGCGGCTGGACCTCTCCGATGAGAAC-3'
rvs- <i>mdg-trunc3</i> (GH5)	5'- <u>G</u> GTGCTCGAGTGCGGCGGGCAATGGCGTTCCAG-3'

### 2.2 Purification of the GH5 and GH5-12 protein

Since an affinity chromatography by Ni-TED column revealed no binding of the His-tagged GH5 and GH5-12 proteins an alternative purification protocol via fractionated ammonium sulfate (AS) precipitation, ion exchange chromatography and size exclusion chromatography was established. For fractionated AS precipitation the protein fraction after heat precipitation was first incubated (30 min, stirring at 4°C) in the presence of 1 M AS and after centrifugation (26000 x g for 30 min at 4°C) the GH5 protein was precipitated with 2.2 M AS and the GH5-12 protein by addition of 2.6 M AS followed by centrifugation (26000 x g for 30 min at 4°C). After resuspension and dialysis against 20 mM TRIS HCl, pH<sup>25°C</sup> 7.0 the protein fractions were applied on a Resource-Q column (6 mL column volume, flow rate 2 mL min<sup>-1</sup>, GE Healthcare) equilibrated in 20 mM TRIS HCl pH<sup>25°C</sup> 7.0, followed by washing with equilibration buffer (20-fold column volume) and elution with a linear salt gradient (0-1 M NaCl, 20-fold column volume, flow rate 2 mL min<sup>-1</sup>, Äkta purifier, GE Healthcare). Fractions containing the respective proteins (checked by SDS-PAGE and activity tests) were pooled and concentrated via Vivaspin concentrators (30 kDa MW cut off, Sartorius Stedim Biotech). After dialysis overnight against 50 mM TRIS HCl, 300 mM NaCl, pH<sup>25°C</sup> 7, both proteins were purified via size exclusion chromatography (Superdex™ 200, HiLoad™ 26/60 column, 320 mL volume, GE Healthcare) equilibrated in dialysis buffer (flow rate 2 mL min<sup>-1</sup>, Äkta purifier, GE Healthcare). Protein fractions containing the purified protein (checked by SDS-PAGE and activity tests) were pooled and concentrated via Vivaspin concentrators (30 kDa MW cut off, Sartorius Stedim Biotech). The protein solutions were stored at -70°C (in the presence of 20% glycerol) and remained active for at least 5 month.

### 3. Results

**Supplementary Table 3.** Glycoside hydrolases (GHs) and carbohydrate esterases (CEs) encoding sequences, identified in the genome of *Thermococcus* strain 2319x1.

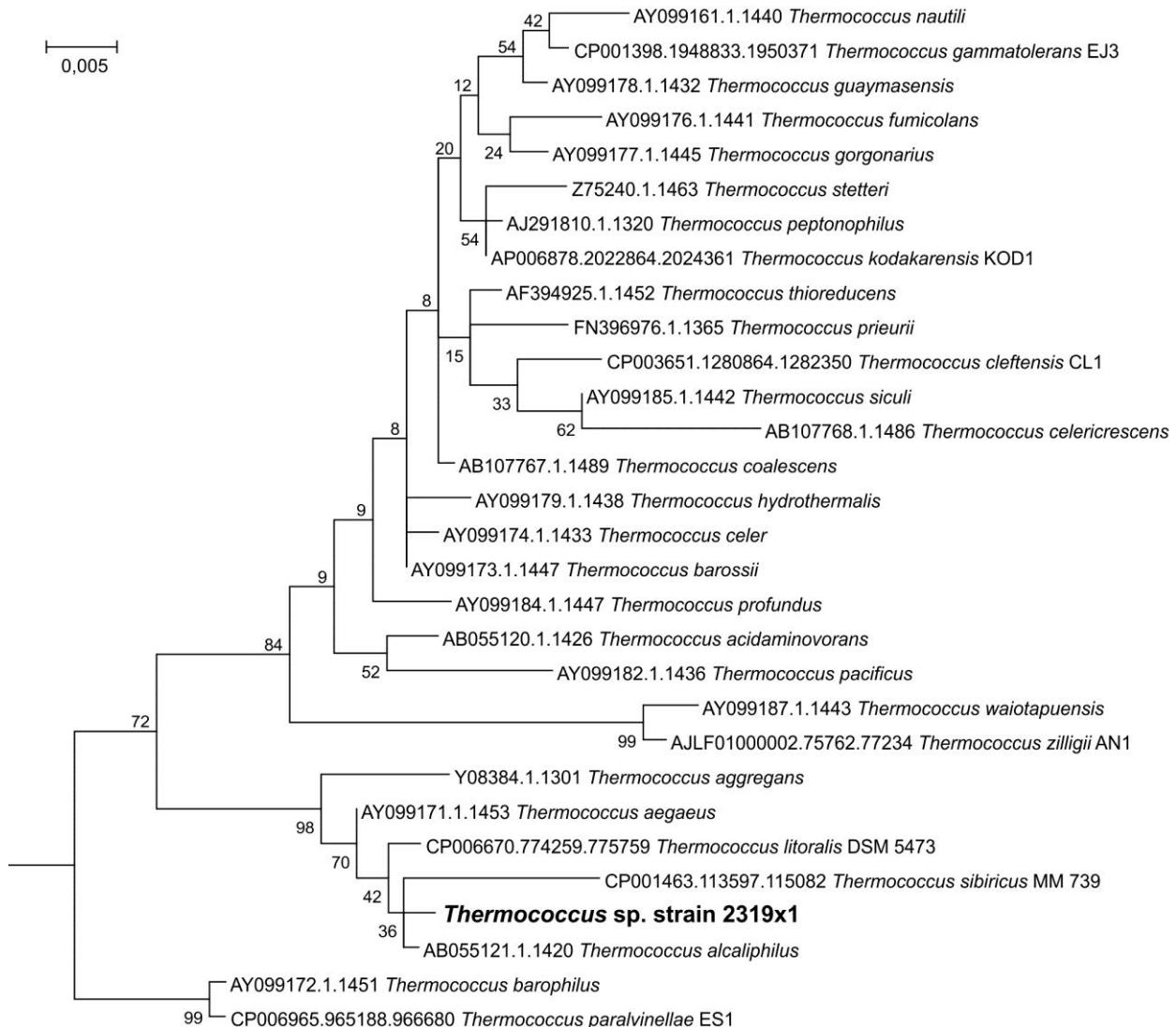
Nº	GenBank ID	Cazy family	Pred Hel=	Topol ogy=	SignalP (G(+)& eukar)
1.	ADU37_CD S00170	Glycoside hydrolase 13 family enzyme, alpha-amylase	0		N
2.	ADU37_CD S02890	Glycoside hydrolase 63 family enzyme, glycogen debranching enzyme, putative	0		N
3.	ADU37_CD S02900	Glycoside hydrolase 130 family enzyme, beta-1,4 mannooligosaccharide phosphorilase	0		N
4.	ADU37_CD S03850	Glycoside hydrolase 57 family enzyme, alpha-amylase/4alpha-gluconotransphosphorylase, putative	0		N
5.	ADU37_CD S05320	Glycoside hydrolase 1 family enzyme, beta-glucosidase, putative	0		N
6.	ADU37_CD S05340	Glycoside hydrolase 57 family enzyme	0		N
7.	ADU37_CD S07270	Glycoside hydrolase 35 family enzyme, exo-beta D-glucosaminidase, putative	0		N
8.	ADU37_CD S07340	Glycoside hydrolase 1 family enzyme, beta-galactosidase, putative	0		N
9.	ADU37_CD S07440	Glycoside hydrolase 1 family enzyme, beta-galactosidase, putative	0		N
10.	ADU37_CD S07750	Glycoside hydrolase 122 family enzyme, alpha-glucosidase, putative	0		N
11.	ADU37_CD S12970	Glycoside hydrolase 57 family enzyme, 1,4alpha-branched enzyme	0		N
12.	ADU37_CD S13420	Carbohydrate esterase, putative	0		Y
13.	ADU37_CD S18940	Glycoside hydrolase 57 family enzyme, amylopullulanase, putative	1	o1075- 1097i	Y
14.	ADU37_CD S18980	Glycoside hydrolase 13 family enzyme, neopullulanase, putative	0		N
15.	ADU37_CD S20530	Glycoside hydrolase 13 family enzyme	0		N
16.	ADU37_CD S21030	Glycoside hydrolase 13 family enzyme	1	i7-25o	Y
17.	ADU37_CD S21840	Putative oxidoreductase, related to GH109 family	0		N
18.	ADU37_CD S22600	Multidomain glycoside hydrolase protein, endoglucanase/endoxylanase	1	i7-29o	Y

PredHel and Topology= are results of transmembrane regions predictions by the TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM>). "iXXX-oYYY" indicates the localization and coordinates of the transmembrane (TM) region in the protein with „i“ inside and “o” outside. TM predictions at the N-terminal end of the protein often indicate the presence of a signal peptide (i.e. i7-25o in ADU37\_CDS21030, i7-29o in ADU37\_CDS22600). For ADU37\_CDS18940 a C-terminal TM is predicted (protein region o1075-1097i).

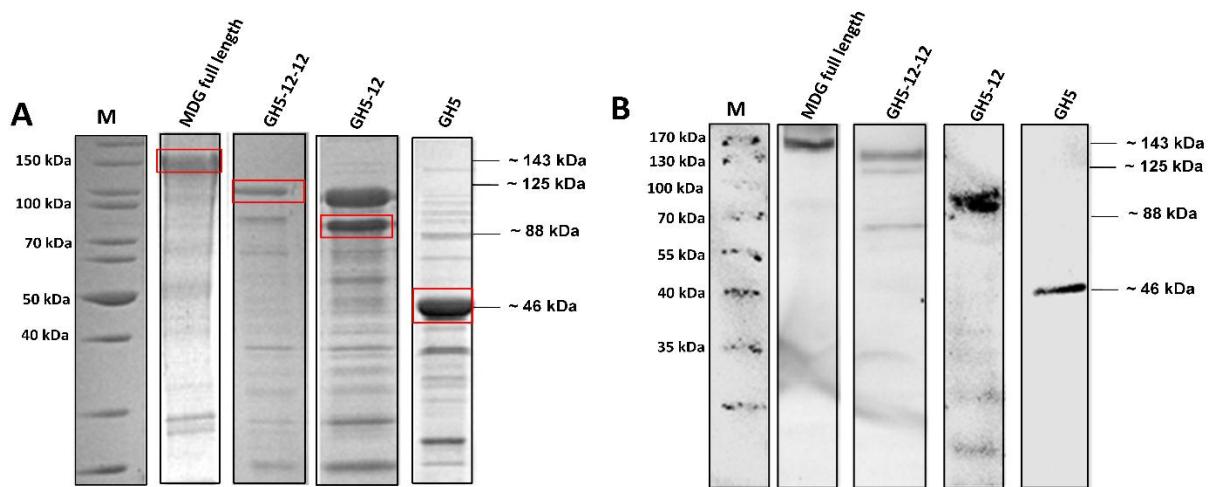
SignalP (G(+)& eukar): results of signal peptide predictions by the SignalP server (for Gram-positives and Eukaryotes; <http://www.cbs.dtu.dk/services/SignalP/index.php>).



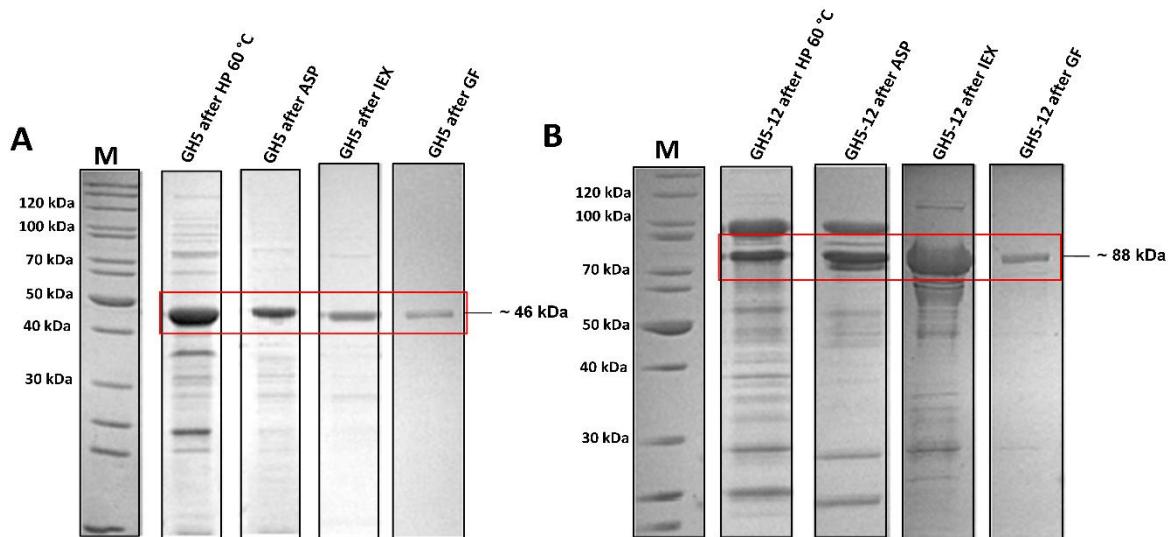
**Supplementary Figure 2.** Sampling site and *in situ* enrichment cultures. Arrows indicate the 50 mL bottles and 18 mL Hungate tubes incubated in the hot spring, located in the tidal zone near Goryachiy cape of Kunashir Island (South Kurils, Russian Far East region).



**Supplementary Figure 3.** 16S rRNA gene sequence-based Maximum likelihood phylogenetic tree of *Thermococcus* sp. strain 2319x1 and validly published representatives of the *Thermococcus* genus. The tree with the highest log likelihood (-3041.4022) is shown. The number of trees of totally 100 repetitions (bootstrap analysis) in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 31 nucleotide sequences. All positions with less than 95% site coverage were eliminated. There were a total of 1286 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). Bar, 0.50 nucleotides substitutions per 100. *Thermococcus* sp. strain 2319x1 is shown in bold. *Pyrococcus yayanosii* strain CH1 (NR 102853.1) was used as outgroup.



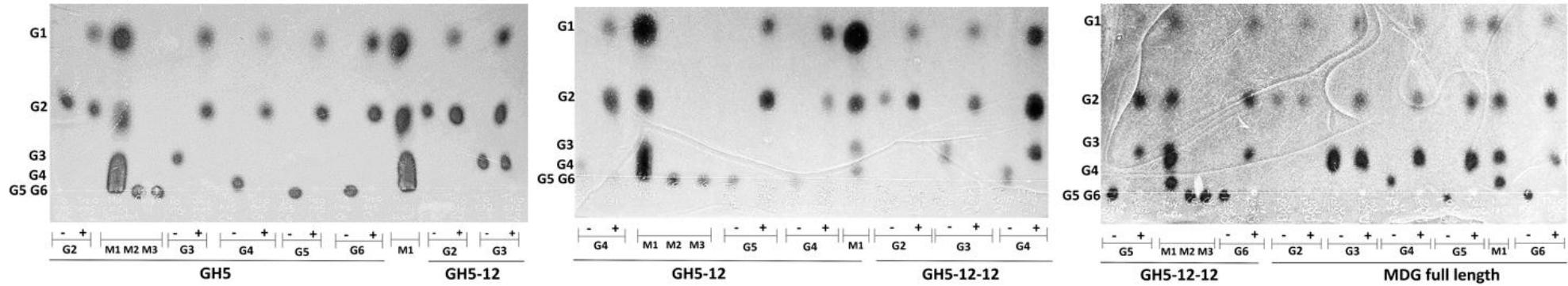
**Supplementary Figure 4.** Expression of the recombinant MDG and the different truncated versions in *Escherichia coli*. (A) Protein fractions (5 µg protein each lane) after heat precipitation at 60°C (30 min) were separated via SDS-PAGE (12.5% SDS-PAGE, coomassie staining) and (B) expressed proteins were identified by immunodetection via anti His-tag antibodies. Marker (M): unstained (A) and prestained (B) protein standard (Fermentas); MDG full length (~ 143 kDa); GH5-12-12 (~ 125 kDa); GH5-12 (~ 88 kDa); single GH5 (~ 46 kDa))



**Supplementary Figure 5.** Purification of the truncated GH5 (A) and GH5-12 (B) protein. Purification was followed by SDS-PAGE (12.5% SDS-PAGE, coomassie staining). Protein fractions (5 µg protein) after heat precipitation (HP), ammonium sulphate precipitation (ASP), ion exchange chromatography (IEX) and protein fractions (2 µg protein) after size exclusion chromatography (SEC) are shown. Marker (M) :unstained protein standard (Fermentas).

**Supplementary Table 4.** Substrate specificity and specific activity of the full length MDG and the truncated MDG proteins.

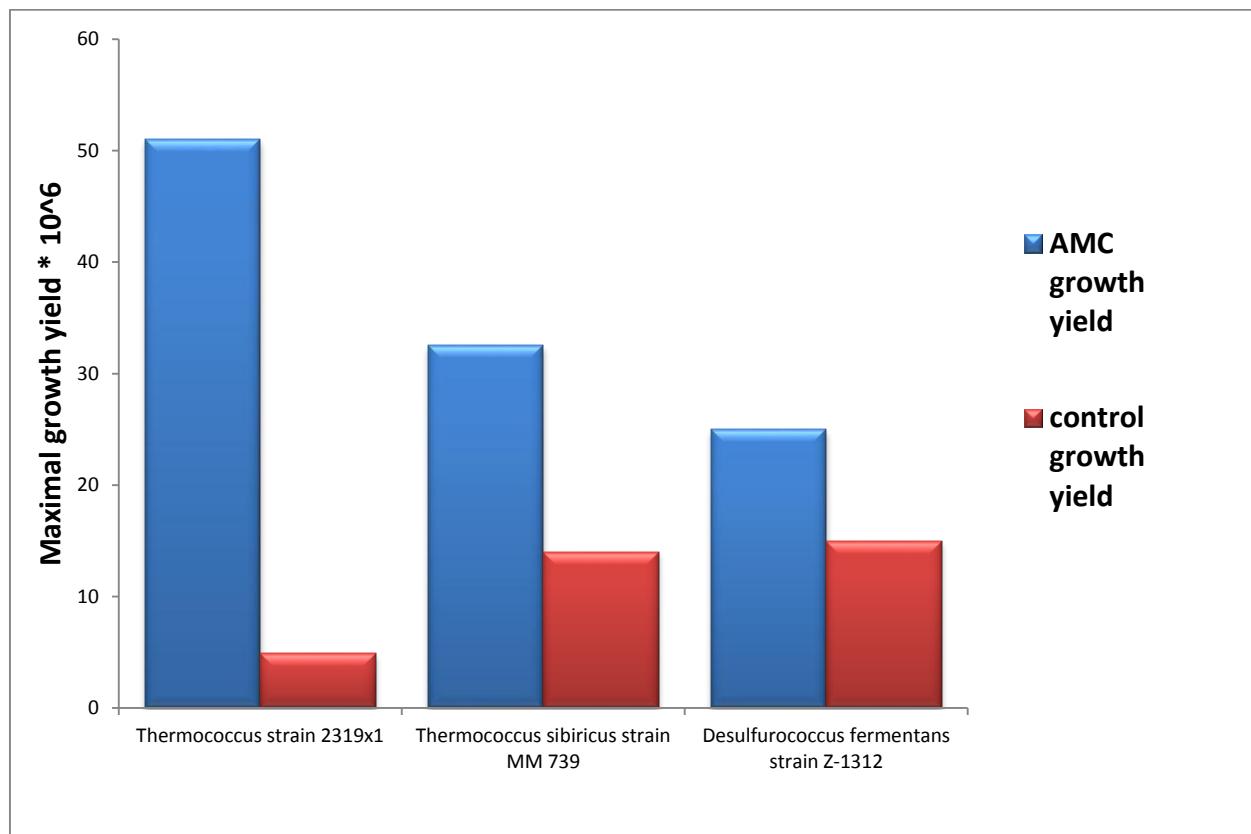
Substrate	Backbone linkage(s)	Specific activities [U mg <sup>-1</sup> ] (K <sub>cat</sub> [s <sup>-1</sup> ])			
		GH5	GH5-12	GH5-12-12	MDG full length
<b>Glucan oligosaccharides</b>					
Cellobiose	β-1,4 only	0.5 (1498)	0.6 (3403)	1.4 (10248)	1.0 (8937)
Cellobiohexaose	β-1,4 only	2.3 (6293)	2.7 (14380)	6.3 (47120)	5.6 (48250)
<b>Glucan polysaccharides</b>					
Barley glucan	β-1,3/4	4.3 (13606)	5.6 (29832)	15.9 (119475)	14.2 (121664)
Lichenan	β-1,3/4	7.01 (19347)	7.02 (37065)	7.44 (39283)	6.75 (57915)
Carboxymethyl cellulose	β-1,4 only	2.7 (7452)	3.4 (17952)	6.7 (50250)	5.9 (50622)
Hydroxyethyl cellulose	β-1,4 only	2.02 (5575)	2.6 (13728)	4.7 (35250)	3.3 (30030)
Xyloglucan	β-1,4/6	0.88 (2428)	0.88 (4646)	0.94 (4963)	0.95 (8151)
MCC Avicel PH101	β-1,4 only	0.8 (2152)	0.8 (4418)	0.8 (5920)	0.7 (5920)
PAT Avicel PH101	β-1,4 only	2.0 (5437)	2.5 (13200)	5.0 (36750)	3.1 (26169)
Curdlan	β-1,3 only	0.39 (1076)	0.39 (2059)	0.5 (3750)	0.38 (3260)
<b>Mannans</b>					
Galactomannan	β-1,4/6	9.2 (25392)	6.3 (33264)	2.09 (15675)	2.2 (18876)
<b>Xylans</b>					
Birchwood xylan	β-1,4 only	0.27 (745)	0.29 (1531)	0.25 (1875)	0.23 (1973)



**Supplementary Figure 6. Degradation of cellobiose and cellooligosaccharides by the MDG and its truncated versions.**

Substrate solutions (1% (w/v)) were mixed with 25 µg of each protein and either directly transferred to ice and the reaction was stopped by the addition of 80 % (v/v) acetone (-), or incubated for 2 h at 60°C (+). Hydrolysis products were separated on aluminum sheet (20 x 20 cm) silica gel 60/kieselguhr F<sub>254</sub> plates (Merck) using a mixture of ethyl acetate, methanol and H<sub>2</sub>O (68:23:9, v/v/v) as mobile phase. Products were visualized with KMnO<sub>4</sub> solution (1.5 g KMnO<sub>4</sub>, 10 g K<sub>2</sub>CO<sub>3</sub> and 1.25 ml 10 % aq. NaOH in 200 ml H<sub>2</sub>O). G: Number of glucose units. G1 = glucose; G2 = cellobiose; G3 = cellotriose; G4 = cellotetraose; G5 = cellopentaose; G6 = cellohexaose. Marker: M1 = G1, G2, G3, G4. M2 = G5. M3 = G6.

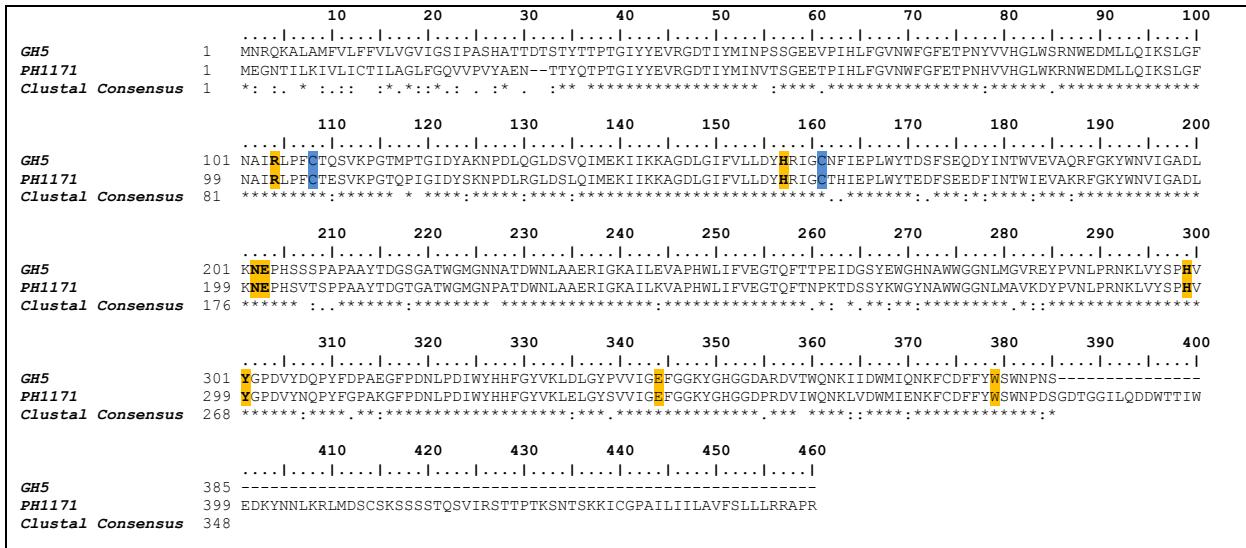
## 4. Discussion



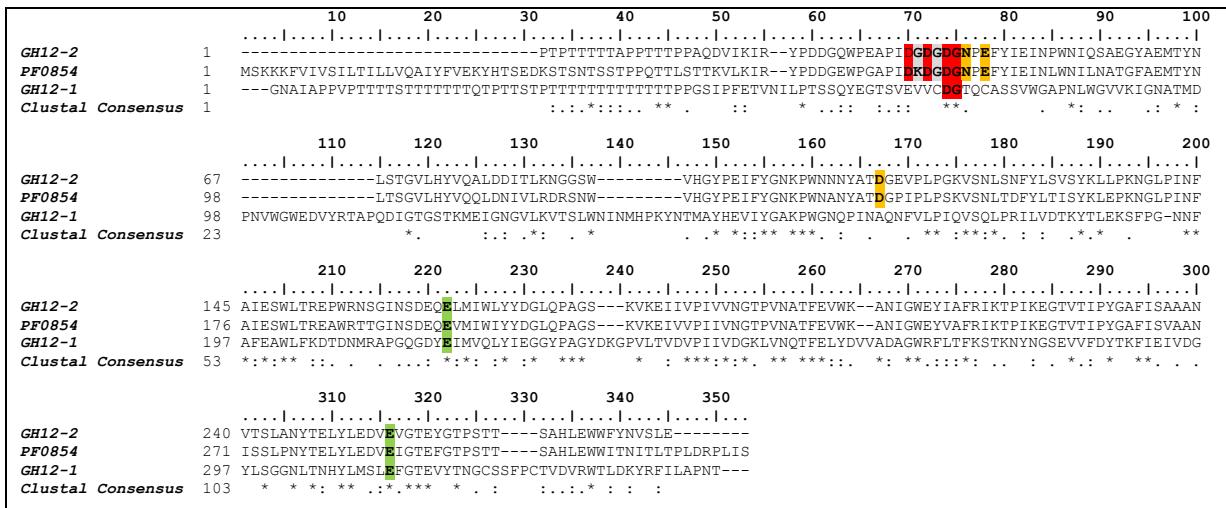
**Supplementary Figure 7.** Comparative analysis of growth yields of three hyperthermophilic archaeal strains, growing on amorphous cellulose (AMC). All strains were cultivated under their optimal growth conditions. Control means the same medium without AMC substrate.

**Supplementary Table 5. Best BLAST hits of CBM2-1 and CBM2-2.**

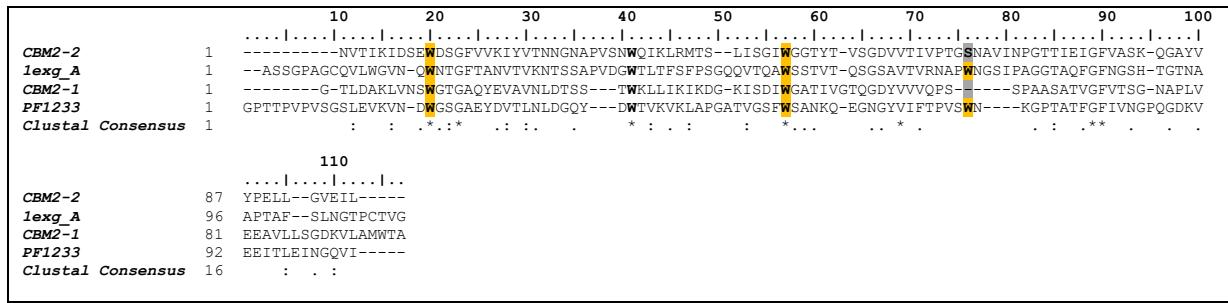
CBM2-1 (1103-1199)		max	total	% query	% identity	e-value	acession
1	Cel12E [unidentified prokaryotic organism]	50,4	50,4	0,97	0,37	0,00006	CRL09147.1
2	chitinase [Pyrococcus furiosus]	50,4	50,4	0,97	0,35	0,00007	WP_011012376.1
3	Chain A, Crystal Structure Of Chitin Biding Domain Of Chitinase From Pyrococcus furiosus	46,6	46,6	0,94	0,35	0,0002	2CWR_A
4	chitinase [Thermococcus kodakarensis]	47	47	0,9	0,34	0,0007	WP_048053867.1
5	glycoside hydrolase family 5 [Cyanobacterium aponinum PCC 10605]	47,4	47,4	0,73	0,35	0,0008	AFZ54640.1
6	chitinase [Thermococcus kodakaraensis]	47	47	0,9	0,34	0,001	BAA88380.1
7	hypothetical protein, partial [Thermococcus nautili]	44,7	44,7	0,9	0,29	0,006	WP_042693178.1
8	Chitinase [Thermococcus nautili]	44,7	44,7	0,9	0,29	0,007	AHL22682.1
9	Glycoside Hydrolase Family 5 candidate endoglucanase [Bifidobacterium thermophilum RBL67]	43,9	43,9	0,72	0,35	0,009	AGH41463.1
CBM2-2 (1207-1303)		max	total	% query	% identity	e-value	acession
1	chitinase [Mycobacterium kansasii]	68,9	135	0,93	0,36	2,00E-11	WP_036448276.1
2	hypothetical protein [Mycobacterium kansasii]	68,9	135	0,93	0,36	4,00E-11	WP_036391652.1
3	hypothetical protein [Mycobacterium kansasii]	68,6	135	0,93	0,36	4,00E-11	WP_042313822.1
4	hypothetical protein [Mycobacterium gastri]	67,4	131	0,93	0,33	1,00E-10	WP_036410849.1
5	hypothetical protein [Mycobacterium kansasii]	66,6	66,6	0,93	0,33	1,00E-10	WP_036444190.1
6	hypothetical protein [Mycobacterium kansasii]	63,2	63,2	0,93	0,34	1,00E-10	WP_042313524.1
7	hypothetical protein MGAST_01715 [Mycobacterium gastri 'Wayne']	67	131	0,93	0,33	2,00E-10	ETW25608.1
8	glycoside hydrolase [Calothrix sp. 336/3]	66,6	66,6	0,91	0,38	2,00E-10	WP_046815061.1
9	cellulose binding domain protein [Mycobacterium kansasii 824]	61,6	61,6	0,89	0,34	3E-09	ETZ98981.1
10	endoglucanase [Microcoleus sp. PCC 7113]	62,8	62,8	0,95	0,32	4E-09	WP_015184298.1
11	hypothetical protein [Mycobacterium kansasii]	62,8	62,8	0,9	0,34	6E-09	WP_023367572.1
12	hypothetical protein [Mycobacterium kansasii]	62,4	62,4	0,9	0,34	6E-09	WP_041327210.1
13	mannan endo-1,4-beta-mannosidase [Streptomyces xiamenensis]	61,2	61,2	0,9	0,37	1E-08	WP_046725132.1
14	chitinase [Mycobacterium kansasii]	61,2	61,2	0,89	0,34	1E-08	KEP39480.1
15	mannan endo-1,4-beta-mannosidase [Streptomyces sp. NRRL F-2890]	61,2	61,2	0,9	0,37	1E-08	WP_030731241.1
16	hypothetical protein MPTA5024_00600 [Microbispora sp. ATCC PTA-5024]	60,8	60,8	0,85	0,35	1E-08	ETK38093.1
17	chitinase [Mycobacterium kansasii]	57,8	57,8	0,9	0,3	1E-08	WP_036394018.1
18	hypothetical protein [Microbispora sp. ATCC PTA-5024]	60,5	60,5	0,88	0,34	2E-08	WP_036322646.1
19	chitinase [Mycobacterium liflandii]	60,8	60,8	0,9	0,31	2E-08	WP_041300546.1
20	hypothetical protein [Calothrix sp. PCC 7103]	60,5	60,5	0,98	0,33	2E-08	WP_035173330.1
21	PE-PGRS family protein [Mycobacterium marinum M]	60,5	118	0,9	0,31	3E-08	ACC42687.1
22	xyloglucanase [Streptomyces griseoflavus]	60,5	60,5	0,95	0,31	3E-08	WP_004922772.1
23	chitinase [Mycobacterium marinum]	60,5	118	0,93	0,31	3E-08	WP_041324889.1
24	hypothetical protein [Mycobacterium gastri]	60,5	60,5	0,93	0,33	4E-08	WP_036414736.1
25	hypothetical protein [Mycobacterium ulcerans]	60,5	117	0,9	0,31	4E-08	WP_011739252.1



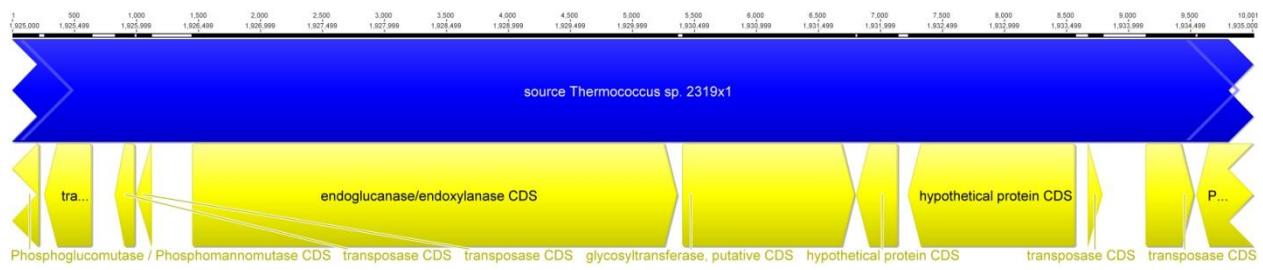
**Supplementary Figure 8.** Sequence Alignment of the MDG GH5 domain (residues 1 - 385) and the endo-1,4- $\beta$ -glucanase PH1171 (EGPh) from *Pyrococcus horikoshii* (Ando et al. 2002; Kashima et al. 2005). All identical amino acid residues are marked with a star (\*). Conserved residues that are similar to the catalytic cleft of PH1171 (PH1171: Arg102, His155, Asn200, Glu201, His297, Tyr299, Glu342 and Trp377) are highlighted in orange. Cysteins that are involved in the formation of a disulfide bond (PH1171: Cys106 – Cys159) are marked in blue. Alignments were performed with ClustalX and edited with BioEdit.



**Supplementary Figure 9.** Sequence Alignment of the MDG GH12-1 domain (residues 416 - 761), the MDG GH12-2 domain (residues 791 - 1070) (both including upstream linker regions) and the well characterized endo-1,4- $\beta$ -glucanase PF0854 (EGPf) of *Pyrococcus furiosus* (Bauer et al. 1999). The nucleophile and the proton donor (PF0854: E197, E290, respectively) conserved in family 12 GHs are found in both GH12 domains of the MDG and are highlighted in green. The calcium-binding motif (DxDxDG) as well as the residues involved in metal ion coordination identified in PF0854 (PF0854: Asp68, Asp70, Asp72, Asn74, Glu76 and Asp142) are found in the upstream linker region of GH12-2 (with low sequence similarity), except Asp142 which is located in the predicted GH12-2 domain. Residues for binding and coordination of Ca<sup>2+</sup> are highlighted in red and orange, respectively. Alignments were performed with ClustalX and edited with BioEdit.



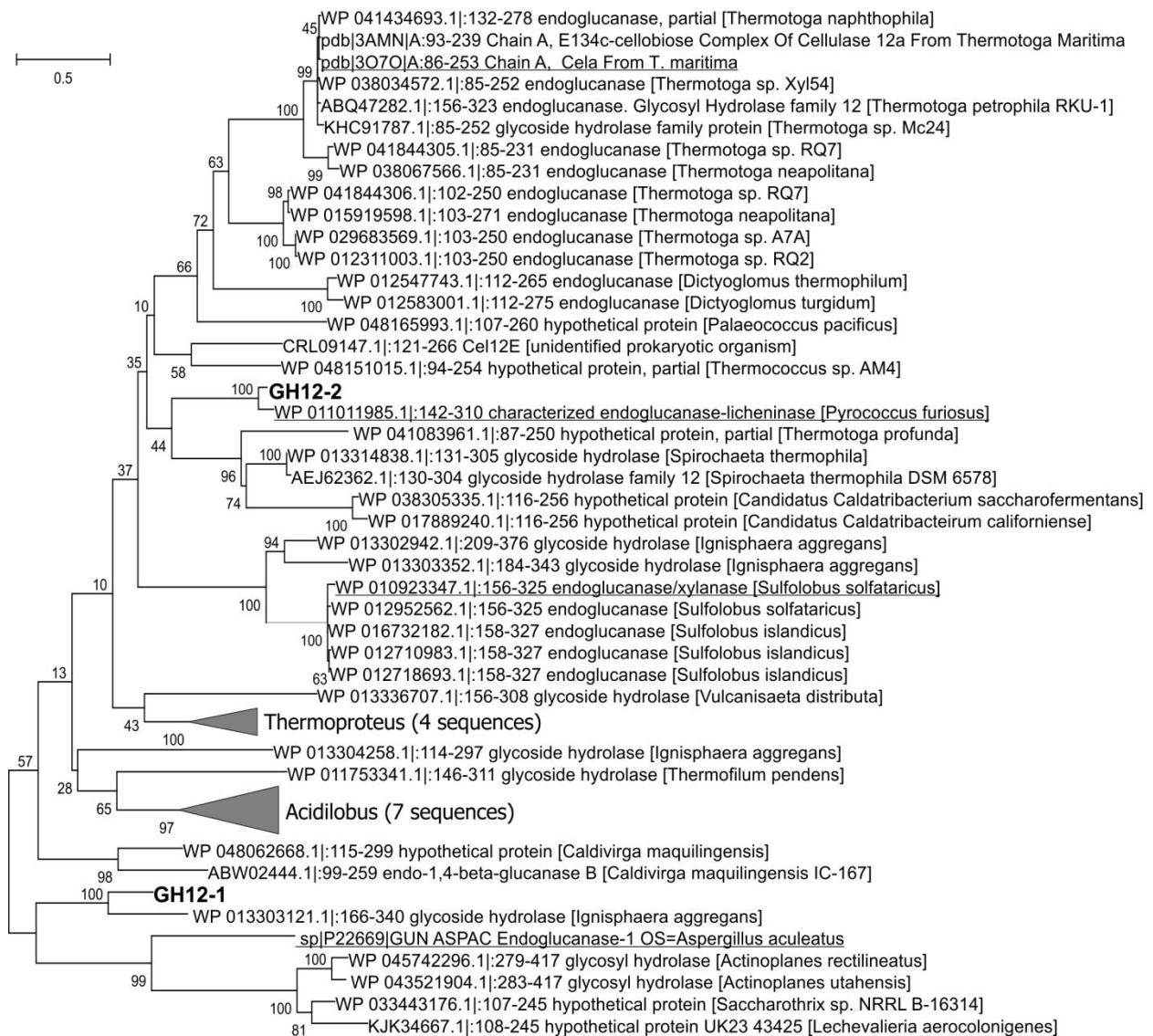
**Supplementary Figure 10.** Sequence Alignment of the MDG CBM2-1 domain (residues 1103 - 1199), the MDG CBM2-2 domain (residues 1208 - 1303), 1exg\_A of *Cellulomonas fimi* and PF1233 of *Pyrococcus furiosus* (Xu et al. 1995; Nakamura et al. 2008; Simpson et al. 2000). Conserved surface-exposed tryptophans in CBM domains are highlighted in orange. The third conserved tryptophan residue is only present in 1exg\_A of *Cellulomonas fimi* and PF1233 of *Pyrococcus furiosus*, whereas it is completely missing in the CBM2-1 domain and replaced by a serine in the CBM-2 domain of the MDG (differences are highlighted in grey). Alignments were performed with ClustalX and edited in BioEdit.



**Supplementary Figure 11.** The *mdg* gene is encoded in a gene cluster together with a putative glycosyltransferase and two hypothetical proteins in the genome of *Thermococcus* sp. strain 2319x1. The gene cluster is flanked by transposase encoding genes.



**Supplementary Figure 12.** Maximum likelihood phylogenetic tree of the GH5 domain of the novel multidomain glycosidase (MDG, ADU37\_CDS22600) and the nearest relatives. The tree with the highest log likelihood (-9647.9039) is shown. The number of trees of totally 100 repetitions (bootstrap analysis) in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 40 amino acid sequences. All positions with less than 95% site coverage were eliminated. There were a total of 292 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). Bar, 20 amino acid substitutions per 100. The *Thermococcus* sp. strain 2319x1 MDG GH5 domain is depicted in bold. The nearest characterized realtive is underlined.



**Supplementary Figure 13.** Maximum likelihood phylogenetic tree of the GH12 domains of the novel multifunctional, multdomain glycosidase (MDG, ADU37\_CDS22600) and the nearest relatives. The tree with the highest log likelihood (-9585.6354) is shown. The number of trees of totally 100 repetitions (bootstrap analysis) in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 54 amino acid sequences. All positions with less than 95% site coverage were eliminated. There were a total of 125 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). Bar, 50 amino acid substitutions per 100. The *Thermococcus* sp. strain 2319x1 MDG GH12-1 and GH12-2 domains are shown in bold. The nearest characterized relatives are underlined.

### Supplementary Literature

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.